Wavelet and Isotonic Regression

Hong-ye Gao
Statistical Sciences Division
MathSoft, Inc.
1700 Westlake Ave. N., Suite 500
Seattle, WA 98109

Abstract

Consider the model:

$$y_i = f(t_i) + z_i$$

where f is a decreasing function and $\{z_i\}$ are assumed to be a stationary Gaussian process with mean zero and variance σ^2. We propose a simple thresholding procedure based on the fact that the wavelet coefficients for f, under Haar basis, are non-negative. We show that our estimator is competitive with the Grenander estimator both theoretically and numerically (in the sense of mean-square-error).

Key Words and Phrases: Isotonic Regression; Monotone Curve; Grenander Estimator; Orthogonal Wavelet Transformation; Shrinkage Estimator.

Acknowledgements: This work was partially supported by ONR-N00014-93-C-0106. The author is grateful to Professor David Donoho for his encouragements, references and helpful comments.

OFFICE OF THE UNDER SECRETARY OF DEFENSE (ACQUISITION)
DEFENSE TECHNICAL INFORMATION CENTER
CAMERON STATION
ALEXANDRIA, VIRGINIA 22304-6145

Jan 25, 1995

IN REPLY REFER TO DTIC-OCC

SUBJECT: Distribution Statements on Technical Documents

TO:
Office of the Chief of Naval Research
800 north Quincy Street
Arlington, VA 22217-5000
Code 22

1995 0925 075

1. Reference: DoD Directive 5230.24, Distribution Statements on Technical Documents, 18 Mar 87.

2. The Defense Technical Information Center received the enclosed report (referenced below) which is not marked in accordance with the above reference.

Wavelet and Isotonic Regression
N00014-93-C-0106

3. We request the appropriate distribution statement be assigned and the report returned to DTIC within 5 working days.

4. Approved distribution statements are listed on the reverse of this letter. If you have any questions regarding these statements, call DTIC's Cataloging Branch, (703) 274-6837.

FOR THE ADMINISTRATOR:

1 Encl

GOPALAKRISHNAN NAIR
Chief, Cataloging Branch

FL-171
Jul 93

DISTRIBUTION STATEMENT A:

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION IS UNLIMITED

DISTRIBUTION STATEMENT B:

DISTRIBUTION AUTHORIZED TO U.S. GOVERNMENT AGENCIES ONLY; (Indicate Reason and Date Below). OTHER REQUESTS FOR THIS DOCUMENT SHALL BE REFERRED TO (Indicate Controlling DoD Office Below).

DISTRIBUTION STATEMENT C:

DISTRIBUTION AUTHORIZED TO U.S. GOVERNMENT AGENCIES AND THEIR CONTRACTORS; (Indicate Reason and Date Below). OTHER REQUESTS FOR THIS DOCUMENT SHALL BE REFERRED TO (Indicate Controlling DoD Office Below).

DISTRIBUTION STATEMENT D:

DISTRIBUTION AUTHORIZED TO DOD AND U.S. DOD CONTRACTORS ONLY; (Indicate Reason and Date Below). OTHER REQUESTS SHALL BE REFERRED TO (Indicate Controlling DoD Office Below).

DISTRIBUTION STATEMENT E:

DISTRIBUTION AUTHORIZED TO DOD COMPONENTS ONLY; (Indicate Reason and Date Below). OTHER REQUESTS SHALL BE REFERRED TO (Indicate Controlling DoD Office Below).

DISTRIBUTION STATEMENT F:

FURTHER DISSEMINATION ONLY AS DIRECTED BY (Indicate Controlling DoD Office and Date Below) or HIGHER DOD AUTHORITY.

DISTRIBUTION STATEMENT X:

DISTRIBUTION AUTHORIZED TO U.S. GOVERNMENT AGENCIES AND PRIVATE INDIVIDUALS OR ENTERPRISES ELIGIBLE TO OBTAIN EXPORT-CONTROLLED TECHNICAL DATA IN ACCORDANCE WITH DOD DIRECTIVE 5230.25, WITHHOLDING OF UNCLASSIFIED TECHNICAL DATA FROM PUBLIC DISCLOSURE, 6 Nov 1984 (Indicate date of determination). CONTROLLING DOD OFFICE IS (Indicate Controlling DoD Office).

The cited documents has been reviewed by competent authority and the following distribution statement is hereby authorized.

A
(Statement)

OFFICE OF NAVAL RESEARCH
CORPORATE PROGRAMS DIVISION
ONR 353
800 NORTH QUINCY STREET
ARLINGTON, VA 22217-5660
(Controlling DoD Office Name)

(Reason)

(Controlling DoD Office Address, City, State, Zip)

Debra T. Hughes
(Signature & Typed Name)

DEBRA T. HUGHES
DEPUTY DIRECTOR
CORPORATE PROGRAMS OFFICE
(Assigning Office)

18 SEP 1995
(Date Statement Assigned)

1 Introduction

Consider the non-parametric regression model:

$$X_i = f(t_i) + z_i \qquad i = 1, ..., n \tag{1}$$

where f is only known to be a decreasing function, $t_i = i/n$ and $\{z_i\}$ are assumed to be a stationary Gaussian process with mean zero and variance σ^2. This kind of regression problem is referred as *isotonic* regression. Examples and interpretations of such models can be found in Barlow *et al* (1972).

Let $f_i = f(t_i)$ and $f = (f_1, ..., f_n)'$ be the sampled version of f. For simplicity we assume that $n = 2^m$ is dyadic. Our goal is to find an estimate $\hat{f}$ depending only on $X_1, ..., X_n$ with small mean-squared-error:

$$R(\hat{f}, f) \equiv n^{-1} \sum_{i=1}^{n} E(\hat{f}_i - f_i)^2$$

In this paper, we will discuss the following 3-step wavelet shrinkage procedure, introduced by Donoho and Johnstone (1992a, b), for estimation of f:

[1] Take the Haar (the simplest wavelet) transform of X_i's to get the empirical wavelet coefficients, $\{y_{j,k}\}_{j=0,...,m-1,k=0,...,2^j-1}$.

Let $\{a_{j,k}\}_{j=0,1,...,k=0,...,2^j-1}$ be the wavelet coefficients of f, from Lemma 1 below, $a_{j,k} \geq 0$ for all j, k.

[2] Apply the threshold:

$$\hat{a}_{j,k} = \begin{cases} 0 & \text{if } y_{j,k} \leq \lambda_j \\ y_{j,k} - \lambda_j & \text{if } y_{j,k} > \lambda_j \end{cases} \tag{2}$$

to the empirical wavelet coefficients $\{y_{j,k}\}$, with some optimally chosen threshlod λ_j, usually $\lambda_j = \sigma\sqrt{2\log(n)/n}$ for all levels.

[3] Invert the Haar tranform to get an estimate $\hat{f}_i$ of f_i.

We will discuss some theoretical properties of the estimator in section 2. In section 3, we will compare our estimator with the Grenander estimate both theoretically and numerically.

2 Theoretical Results

The Haar basis is an orthonormal basis of $L_2[0,1]$ and it is also an interpolating wavelet basis (Donoho, 1992), therefore the wavelet coefficients for the sampled version is essentially the same as those of function version.

We first mention some results on the function version coefficients. Let

$$\phi(t) = I_{[0,1)}(t), \qquad \psi(t) = I_{[0,1/2)}(t) - I_{[1/2,1)}(t)$$

and define

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k)$$

For $f \in L_2[0,1]$, define the wavelet coefficients

$$b_0 = \int f\phi, \qquad a_{j,k} = \int f\psi_{j,k} \tag{3}$$

and for $0 \le t \le 1$,

$$f(t) = b_0 + \sum_{j\ge 0}\sum_{k=0}^{2^j-1} a_{j,k}\psi_{j,k}(t)$$

(in the sense of L_2). Moreover, there is the extremely useful Parseval relation:

$$||\hat{f} - f||^2_{L_2[0,1]} = (\hat{b}_0 - b_0)^2 + \sum_{j,k}(\hat{a}_{j,k} - a_{j,k})^2$$

Consider the class of decreasing functions on $[0,1]$:

$$\mathcal{D}(C) = \{f : f \text{ decreasing and } f(0) - f(1) \le C\}$$

Lemma 1 *Suppose $f \in \mathcal{D}(C)$ and $a_{j,k}$'s are defined as in (3), then*

$$a_{j,k} \ge 0$$

for all $j \ge 0$ and $0 \le k < 2^j$. And furthermore,

$$2^{j/2}\sum_{k=0}^{2^j-1} a_{j,k} \le C/2$$

for all $j \ge 0$.

Proof: For $j \geq 0$ and $k = 0, 1, ..., 2^j - 1$,

$$\begin{aligned} a_{j,k} &= \int_0^1 f(x)\psi_{j,k}(x)dx = 2^{-j/2}\int_0^1 f(\frac{y+k}{2^j})\psi(y)dy \\ &= 2^{-(j+2)/2}\int_0^1 \{f(\frac{y+2k}{2^{j+1}}) - f(\frac{y+2k+1}{2^{j+1}})\}dy \qquad (4) \\ &\leq 2^{-j/2}(f(\frac{k}{2^j}) - f(\frac{k+1}{2^j}))\frac{1}{2} \end{aligned}$$

The non-negativity of $a_{j,k}$ is obvious from (4) and

$$2^{j/2}\sum_{k=0}^{2^j-1} a_{j,k} \leq \frac{1}{2}\sum_{k=0}^{2^j-1}(f(\frac{k}{2^j}) - f(\frac{k+1}{2^j})) = \frac{f(0)-f(1)}{2} \leq C/2$$

This completes the proof.

Next Lemma gives the optimal rate under L_2 norm:

Lemma 2 (Minimax L_2 Risk)

$$\inf_{\hat{f}} \sup_{f \in \mathcal{D}(C)} R(\hat{f}, f) \asymp n^{-2/3}$$

Let

$$\hat{f}_w(t) = \hat{b}_0 + \sum_{j=0}^{m-1}\sum_{k=0}^{2^j-1} \hat{a}_{j,k}\psi_{j,k}(t)$$

be the wavelet estimate of f, as we described earlier, the following Theorem says that $\hat{f}_w$ enjoys the optimal (or near optimal) rate of convergence.

Theorem 1

$$\inf_{\{\lambda_j\}} \sup_{f \in \mathcal{D}(C)} R(\hat{f}_w, f) \asymp n^{-2/3}$$

For $\lambda_j \equiv \sigma\sqrt{2\log(n)/n}$,

$$\sup_{f \in \mathcal{D}(C)} R(\hat{f}_w, f) \leq C(\frac{\log n}{n})^{2/3}(1 + o(1))$$

The proof is in the Appendix.

Remark: In practice, σ^2, the noise level of the data, is usually unknown. It is natural to find an estimator $\hat{\sigma}$, say, to replace the unknown σ in the thresholds. In this paper, $\hat{\sigma}$ is derived from a linear estimate of f

$$\hat{\sigma}^2 = \sum_{j=[m/2]}^{m-1} \sum_{k=0}^{2^j-1} y_{j,k}^2 \tag{5}$$

Since $y_{j,k}$ are independently $N(a_{j,k},\ \sigma^2/n)$ distributed,

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-\sqrt{n}}(\delta_n)$$

is non-central χ^2 distributed with the non-centrality

$$\delta_n^2 = \sum_{j=[m/2]}^{m-1} \sum_{k=0}^{2^j-1} \frac{na_{j,k}^2}{\sigma^2}$$

and for $f \in \mathcal{D}(C)$, from Lemma 1 and Jensen's inequality,

$$\delta_n^2 \le \frac{n}{\sigma^2}\Big(\sum_{j=[m/2]}^{m-1} \sum_{k=0}^{2^j-1} |a_{j,k}|\Big)^2 \le \frac{n}{\sigma^2}\Big(\sum_{j=[m/2]}^{\infty} \frac{C}{2^{j/2}}\Big)^2 = \frac{2C^2\sqrt{n}}{\sigma^2(\sqrt{2}-1)^2}$$

By calculating the first and second momemts of the non-central χ^2 distribution, for $X \sim \chi^2_n(\delta)$, $EX = n + \delta^2$ and $Var(X) = 2n + 4\delta^2$, we have

$$E(\hat{\sigma}^2 - \sigma^2)^2 = O(n^{-1})$$

Finally, by using the inequality

$$(\sqrt{x} - \sqrt{y})^2 \le \frac{(x-y)^2}{y}$$

for all $x \ge 0$ and (fixed) $y > 0$, we have

$$E(\hat{\sigma} - \sigma)^2 = O(n^{-1})$$

and this leads to:

Theorem 2 *For* $\lambda_j \equiv \hat{\sigma}\sqrt{2\log(n)/n}$,

$$\sup_{f \in \mathcal{D}(C)} R(\hat{f}_w, f) \le C'\Big(\frac{\log n}{n}\Big)^{2/3}(1 + o(1))$$

The proof is also in the Appendix.

3 Comparisons with LS Estimator

Least square method are commonly used in regression analysis. The LS estimate for f is

$$\hat{f} = \arg \min_{f_1 \geq \cdots \geq f_n} \sum_{i=1}^{n} (y_i - f_i)^2 \tag{6}$$

Note that $\{\hat{f}_i\}$ are also the maximum likelihood estimates of $\{f_i\}$ when $\{z_i\}$ are iid Gaussian random variables.

An effective algorithm, Pool-Adjacent-Violators Algorithm, has been described in §1.2 of Barlow *et al* (1972) and the resulting estimator is usually called the least concave majorant estimator or Grenander estimator, ref. Grenander (1956), Prakasa Rao (1983). Properties of Grenander estimator can also be found in Bergé (1989) and Wang (1991). For example:

Lemma 3 *Let $\hat{f}_G$ be the Grenander estimator, then*

$$\sup_{f \in \mathcal{D}(C)} R(\hat{f}_G, f) \asymp n^{-2/3}$$

i.e. Grenander estimator achieves the optimal rate.

We have the following comparison:

- Wavelet estimate equipped with the optimal threshold achieves the optimal rate as the Grenander estimate does; while the estimator with the simple shrinkage rule of using $\lambda_j \equiv \hat{\sigma}\sqrt{2\log(n)/n}$ is within a factor $(\log n)^{2/3}$ of the optimal rate.
- Computationally, wavelet procedure is more efficient than Grenander procedure. In fact, the computational effort of the wavelet procedure is of order $n \log n$ comparing with order n^2 of the Grenander procedure.
- Grenander estimator has boundary effects. In fact, it does not converge at discontinous points, Wang (1991). This drawback can be seen from our simulations, while the wavelet estimates do not have boundary effects.

As a final remark, we should mention here is that wavelet procedure does not guarantee us a decreasing solution. From Lemma 1, non-negativity of the Haar coefficients is only a necessary condition for f to be decreasing.

In the case where a decreasing solution is needed, we can still use wavelet procedure as the first step, since wavelet procedure is a rate-preserve transformation. We can then apply Grenander procedure to the transformed data. This two-step procedure requires less computational effort than applying Grenander procedure to the original data.

Numerical examples (Figure 1, 2, 3) show that the wavelet estimate is also competitive numerically with the Grenander estimate. In our simulations, we compare the L_1 and L_2 losses of both Grenander estimator and our wavelet estimator:

$$\begin{aligned} R1 &= \sum_{i=1}^{n} |\hat{f}(t_i) - f(t_i)| \\ R2 &= \sum_{i=1}^{n} (\hat{f}(t_i) - f(t_i))^2 \end{aligned}$$

4 Appendix

To prove our main theorem, we need to introduce some basic facts of Besov space theory. Let's define Besov Body in sequence space l_p

$$\Theta^s_{p,q}(C) = \{\theta = (\theta_{j,k})_{j\geq 0,\, 0\leq k<2^j} : \sum_{j=0}^{\infty} \{2^{jsq} (\sum_{k=0}^{2^j-1} |\theta_{j,k}|^p)^{q/p}\} \leq C\}$$

for $q < \infty$ and

$$\Theta^s_{p,\infty}(C) = \{\theta = (\theta_{j,k})_{j\geq 0,\, 0\leq k<2^j} : \sup_{j\geq 0}\{2^{jsp} \sum_{k=0}^{2^j-1} |\theta_{j,k}|^p\} \leq C^p\}$$

Proof of Theorem 1: From Lemma 1, it is obvious that

$$\Theta_D(C) \subset \Theta_+ \bigcap \Theta^{1/2}_{1,\infty}(C/2) \tag{7}$$

where $\Theta_+ = \{(\theta_{j,k}) : \theta_{j,k} \geq 0\}$.

Therefore,

$$\begin{aligned} \sup_{f \in \mathcal{D}(C)} E||\hat{f}_w - f||^2_{L_2[0,1]} &= \sup_{a \in \Theta_D(C)} E||\hat{a} - a||^2_{l_2} \\ &\leq \sup_{a \in \Theta_+ \bigcap \Theta^{1/2}_{1,\infty}(C/2)} E||\hat{a} - a||^2_{l_2} \end{aligned}$$

It is well known that (Donoho and Johnstone, 1992a, b)

$$\inf_{\tilde{a}} \sup_{a \in \Theta^{1/2}_{1,\infty}(C/2)} E||\tilde{a} - a||^2_{l_2} \asymp n^{-2/3}$$

and for the shrinkage estimates $\hat{a}$ equipped with the optimal thresholds, we have the same rate

$$\sup_{a \in \Theta^{1/2}_{1,\infty}(C/2)} E||\hat{a} - a||^2_{l_2} \asymp n^{-2/3}$$

for the shrinkage estimate $\hat{a}$ equipped with thresholds $\lambda_j = \sigma\sqrt{2\log(n)/n}$,

$$\sup_{a \in \Theta^{1/2}_{1,\infty}(C/2)} E||\hat{a} - a||^2_{l_2} \leq C'(\frac{\log n}{n})^{2/3}$$

Let $\hat{a}_+$ be the projection of $\hat{a}$ onto $\Theta_+ \bigcap \Theta^{1/2}_{1,\infty}(C/2)$, then we have

$$\sup_{a \in \Theta_+ \bigcap \Theta^{1/2}_{1,\infty}(C/2)} E||\hat{a}_+ - a||^2_{l_2} \asymp n^{-2/3}$$

when the optimal thresholds are used, and

$$\sup_{a \in \Theta_+ \bigcap \Theta^{1/2}_{1,\infty}(C/2)} E||\hat{a}_+ - a||^2_{l_2} \leq C'(\frac{\log n}{n})^{2/3}$$

when thresholds $\lambda_j = \sigma\sqrt{2\log(n)/n}$ are used.

It is obvious that the coordinates in $\hat{a}_+$ have the expressions in (2).

Proof of Theorem 2: The proof follows immediately from Theorem 1 and the following Lemma.

Lemma 4

$$\sum_{j=0}^{m-1}\sum_{k=0}^{2^j-1} |E(\hat{a}_{j,k} - a_{j,k})^2 - E(\hat{a}_{j,k} - a_{j,k})^2| = o(n^{-2/3})$$

Proof: From the definitions of $\hat{a}_{j,k}$ and $\hat{a}_{j,k}$,

$$\begin{aligned}
&|(\hat{a}_{j,k} - a_{j,k})^2 - (\hat{a}_{j,k} - a_{j,k})^2| \\
&\quad\leq \frac{\lambda^2}{n}(\hat{\sigma} - \sigma)^2 + \frac{2\lambda\sigma}{n}|z_{j,k}||\hat{\sigma} - \sigma| I_{[z_{j,k} > \lambda(\frac{\hat{\sigma}\wedge\sigma}{\sigma} - \frac{\sqrt{n}a_{j,k}}{\sigma\lambda})]}
\end{aligned}$$

where $z_{j,k} = \sqrt{n}(w_{j,k} - a_{j,k})/\sigma$. So

$$\begin{aligned}
&|E(\hat{a}_{j,k} - a_{j,k})^2 - E(\hat{a}_{j,k} - a_{j,k})^2| \\
&\quad\leq \frac{\lambda^2}{n}E(\hat{\sigma} - \sigma)^2 + \frac{2\lambda\sigma}{n}E\{|z_{j,k}||\hat{\sigma} - \sigma| I_{[z_{j,k} > \lambda(\frac{\hat{\sigma}\wedge\sigma}{\sigma} - \frac{\sqrt{n}a_{j,k}}{\sigma\lambda})]}\} \\
&\quad\leq \frac{\lambda^2}{n}E(\hat{\sigma}-\sigma)^2 + \frac{2\lambda\sigma}{n}(E|z|^{1/\beta})^{\beta}\sqrt{E(\hat{\sigma}-\sigma)^2}P\{z > \lambda(\frac{\hat{\sigma}}{\sigma}\wedge 1 - \frac{\sqrt{n}a_{j,k}}{\sigma\lambda})\}^{\frac{1}{2}-\beta} \\
&\quad\leq O(\frac{\log n}{n^2}) + O(\frac{\sqrt{\log n}}{n^{3/2}}P\{z > \lambda(\frac{\hat{\sigma}}{\sigma}\wedge 1 - \frac{\sqrt{n}a_{j,k}}{\sigma\lambda})\}^{\frac{1}{2}-\beta})
\end{aligned}$$

where $z \sim N(0,\ 1)$, $0 < \beta < 1/2$.

For those $a_{j,k}$ satisfying

$$\frac{\sqrt{n}a_{j,k}}{\sigma\lambda} \leq r$$

where $0 < r < 1$. Let Z_n be a standardized non-central $\chi^2_{n-\sqrt{n}}(\delta_n)$ variate (i.e. $EZ_n = 0$ and $Var(Z_n) = 1$) and $s \in (r, 1)$,

$$\begin{aligned}
&P\{z > \lambda(\frac{\hat{\sigma}}{\sigma} \wedge 1 - \frac{\sqrt{n}a_{j,k}}{\sigma\lambda})\} \\
&\quad\leq P\{z > \lambda(\frac{\hat{\sigma}}{\sigma} \wedge 1 - r)\} \\
&\quad\leq P\{z > \lambda(\frac{\hat{\sigma}}{\sigma} \wedge 1 - r), \hat{\sigma} \geq s\sigma\} + P\{\hat{\sigma} < s\sigma\} \\
&\quad\leq P\{z > \lambda(s - r)\} + P\{\chi^2_{n-\sqrt{n}}(\delta_n) < s^2 n\} \\
&\quad\leq \frac{\phi(\lambda(s-r))}{\lambda(s-r)} + P\{|Z_n| > \frac{(1-s^2)n + \delta_n^2}{\sqrt{2n + 4\delta_n^2}}\}
\end{aligned}$$

$$
\begin{aligned}
&= O\left(\frac{1}{n^{(s-r)^2}\sqrt{\log n}}\right) + O(n^{-1}) \\
&= O\left(\frac{1}{n^{(s-r)^2}\sqrt{\log n}}\right)
\end{aligned}
$$

Consider

$$N_n = \#\{\frac{\sqrt{n}a_{j,k}}{\sigma\lambda} > r,\ 0 \le j < m,\ 0 \le k < 2^j\}$$

where $m = \log_2 n$. Then for $f \in \mathcal{D}(C)$, from Lemma 1,

$$\frac{\sqrt{2}C}{\sqrt{2}-1} \ge \sum_{j=0}^{m-1} \frac{C}{2^{j/2}} \ge \sum_{j=0}^{m-1}\sum_{k=0}^{2^j-1} a_{j,k} \ge \sum_{a_{j,k} > r\sigma\lambda/\sqrt{n}} a_{j,k} \ge \frac{r\sigma\lambda N_n}{\sqrt{n}}$$

and so

$$N_n \le \frac{\sqrt{2n}C}{(\sqrt{2}-1)r\sigma\lambda} \tag{8}$$

Therefore, by choosing r, β small and s close to 1 such that

$$(s-r)^2(1/2-\beta) > 1/6$$

we have

$$
\begin{aligned}
&\sum_{j=0}^{m-1}\sum_{k=0}^{2^j-1} |E(\hat{a}_{j,k} - a_{j,k})^2 - E(\hat{a}_{j,k} - a_{j,k})^2| \\
&= \sum_{j=0}^{m-1}\sum_{k=0}^{2^j-1} \{O(\frac{\log n}{n^2}) + O(\frac{\sqrt{\log n}}{n^{3/2}} P\{z > \lambda(\frac{\hat{\sigma}}{\sigma}\wedge 1 - \frac{\sqrt{n}a_{j,k}}{\sigma\lambda})\}^{\frac{1}{2}-\beta})\} \\
&= O(\frac{\log n}{n}) + O(\frac{\sqrt{\log n}}{n^{3/2}} \sum_{j=0}^{m-1}\sum_{k=0}^{2^j-1} P\{z > \lambda(\frac{\hat{\sigma}}{\sigma}\wedge 1 - \frac{\sqrt{n}a_{j,k}}{\sigma\lambda})\}^{\frac{1}{2}-\beta}) \\
&= O(\frac{\log n}{n}) + O(\frac{\sqrt{\log n}}{n^{3/2}}\{ \sum_{a_{j,k} \le r\sigma\lambda/\sqrt{n}} \frac{1}{n^{(s-r)^2(1/2-\beta)}\sqrt{\log n}} + \sum_{a_{j,k} > r\sigma\lambda/\sqrt{n}} 1\}) \\
&= O(\frac{\log n}{n}) + O(\frac{\sqrt{\log n}}{n^{3/2}}(\frac{n}{n^{(s-r)^2(1/2-\beta)}\sqrt{\log n}} + \sqrt{\frac{n}{\log n}})) \\
&= o(n^{-2/3})
\end{aligned}
$$

This completes the proof of the Lemma.

References

[1] Barlow, R. E., Bartholomew, D. J., Bremner, J. M. and Brunk, H. D. (1972). *Statistical Inference under Order Restrictions.* Wiley, New York.

[2] Birgé, L. (1989). The Grenander Estimator: A Nonasymptotic Approach. *Ann. Statist.* **17**, 1532-1549.

[3] Donoho, D. (1992). Interpolating Wavelet Transforms. Technical Report, Department of Statistics, Stanford University.

[4] Donoho, D. and Johnstone, I (1992a). Minimax Estimation via Wavelet Shrinkage. Technical Report, Department of Statistics, Stanford University.

[5] Donoho, D. and Johnstone, I (1992b). Adapting to Unknown Smoothness via Wavelet Shrinkage. Technical Report, Department of Statistics, Stanford University.

[6] Grenander, U. (1956). On the Theory of Mortality Measurement, Part II. *Skand. Akt.*, **39**, 125-153.

[7] Prakasa Rao (1983). *Nonparametric Functional Estimation.* Academic Press, New York.

[8] Wang, Yazhen (1991). Contributions to the Theory of Estimation of Monotone and Unimodal Densities. Manuscript.

Figure 1

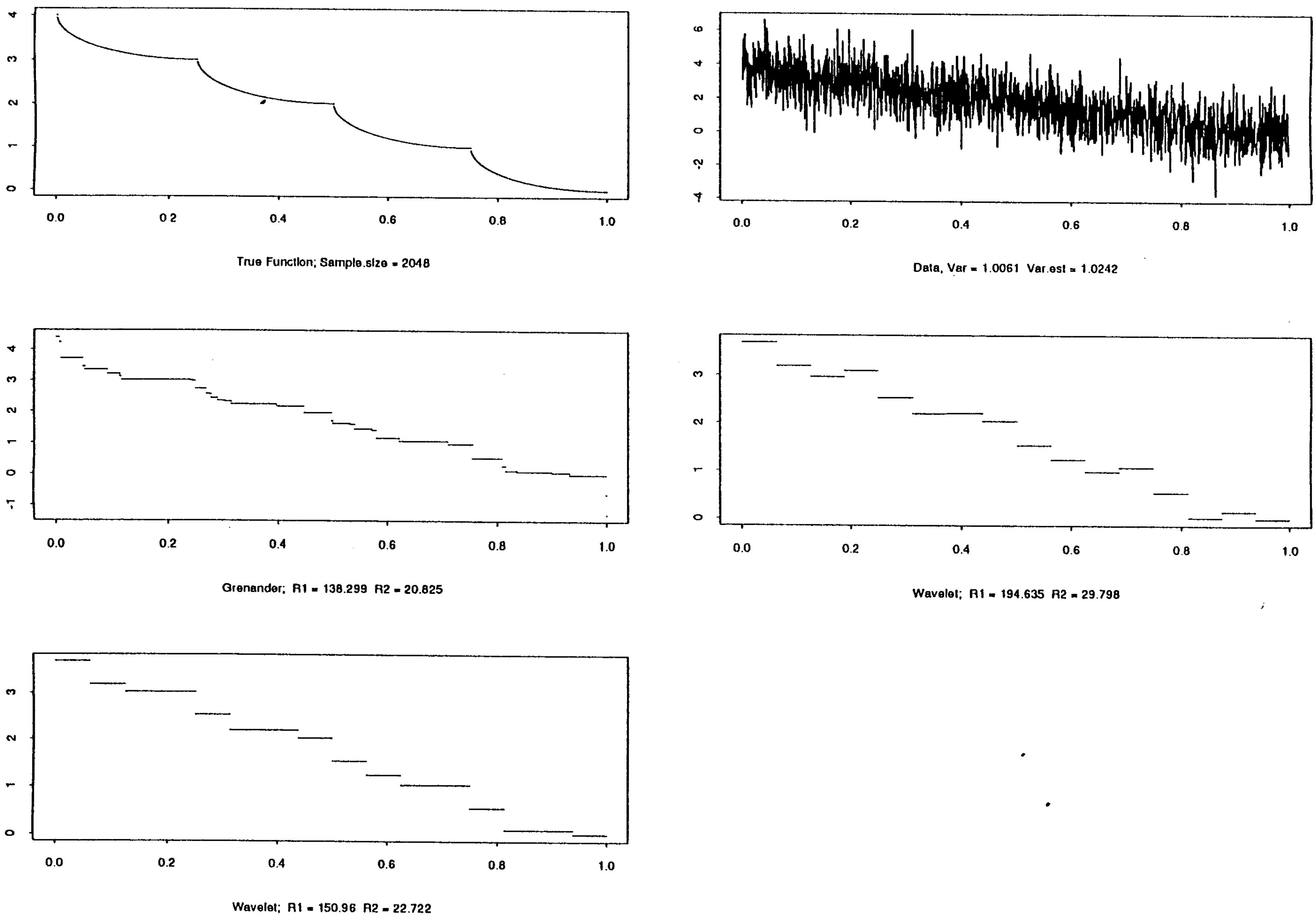
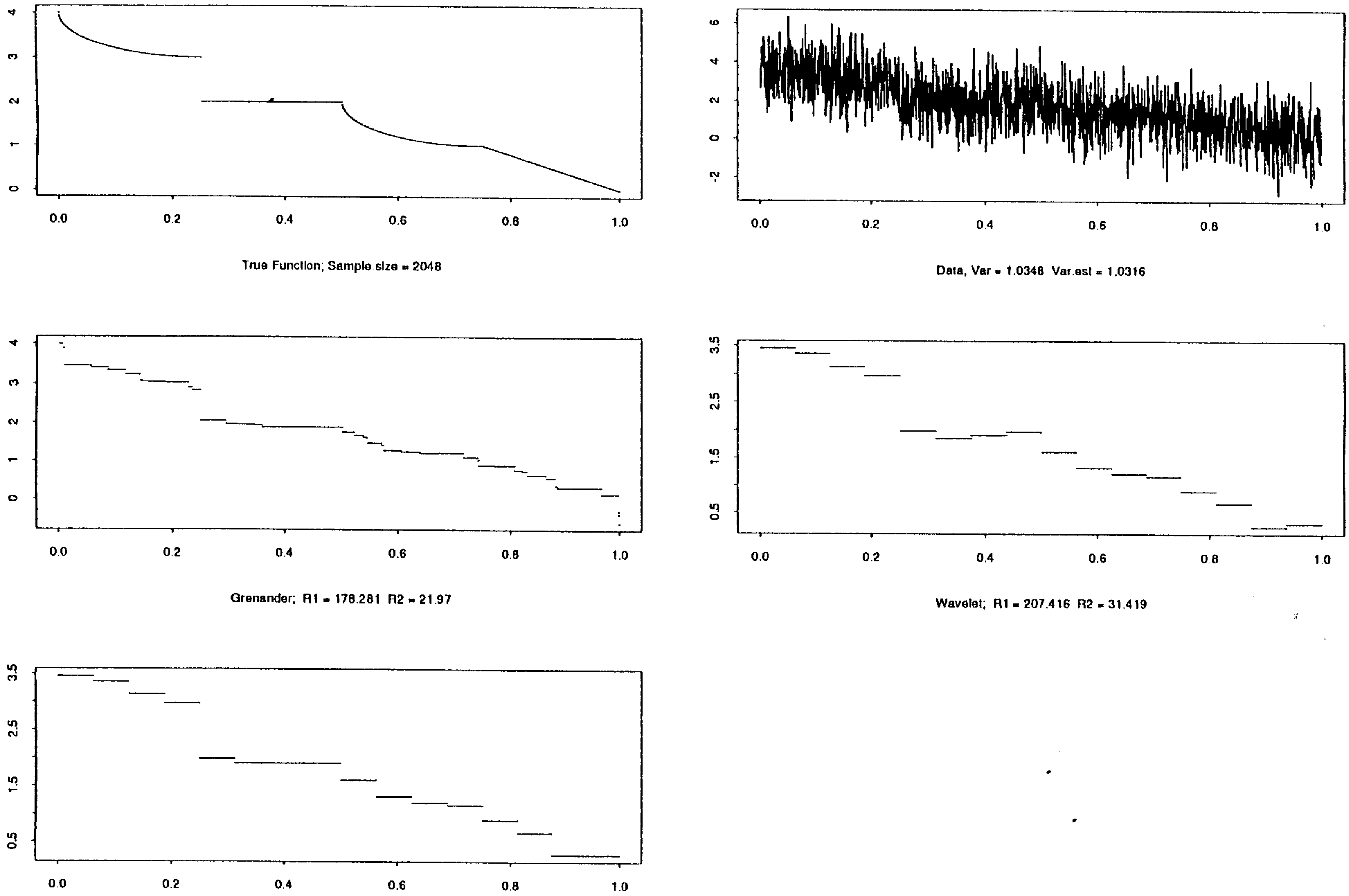

True Function; Sample.size = 2048
Data, Var = 1.0061 Var.est = 1.0242
Grenander; R1 = 138.299 R2 = 20.825
Wavelet; R1 = 194.635 R2 = 29.798
Wavelet; R1 = 150.96 R2 = 22.722

Figure 2



True Function; Sample.size = 2048

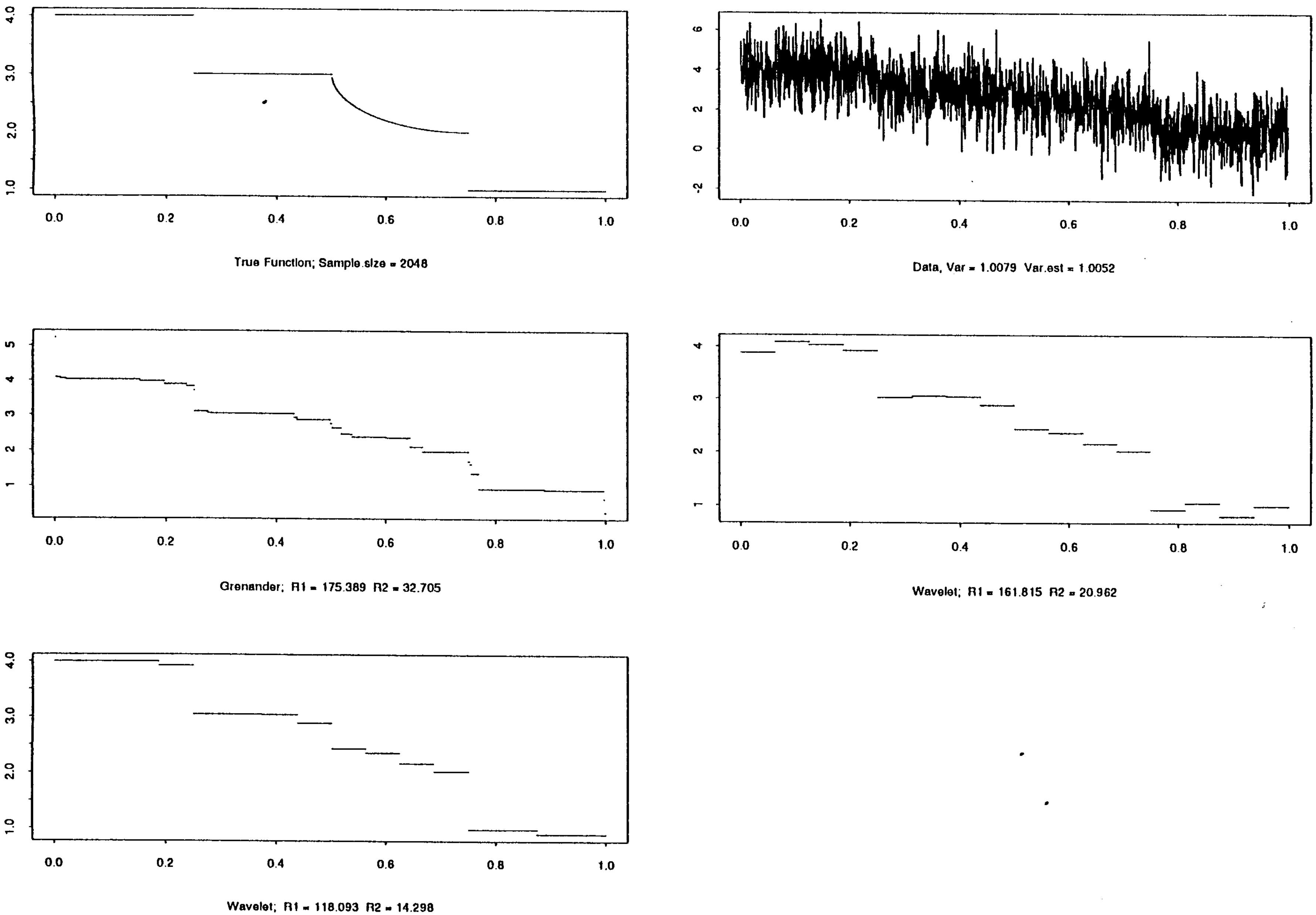
Data, Var = 1.0348 Var.est = 1.0316

Grenander; R1 = 178.281 R2 = 21.97

Wavelet; R1 = 207.416 R2 = 31.419

Wavelet; R1 = 199.963 R2 = 28.384

Figure 3


True Function; Sample.size = 2048
Data, Var = 1.0079 Var.est = 1.0052
Grenander; R1 = 175.389 R2 = 32.705
Wavelet; R1 = 161.815 R2 = 20.962
Wavelet; R1 = 118.093 R2 = 14.298